DOCUMENT RESUME

ED 143 072 EA 009 775

AUTHOR Worthen, Blaine R.; Rogers, W. Todd
TITLE Uses and Abuses of Adversary Evaluation: A Consumer's Guide.
PUB DATE Mar 77
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, N.Y., April 4-8, 1977); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Conflict; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Models; Problem Solving; *Program Evaluation
IDENTIFIERS *Adversary Evaluation

ABSTRACT

The major potentials and pitfalls of adversary evaluation in education are discussed. Reasons why the courtroom model may have limited utility, and the difficulties in the debate model are identified. It is argued that the existence of opposing viewpoints is the core of adversary evaluation, not adherence to existing formats for presenting contrasting views. It is suggested that educational evaluators might develop more appropriate adversary methods tailored specifically for the field of education. Nine issues are addressed that should be considered by anyone intending to use the adversary approach. (Author/MLF)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

BEST COPY AVAILABLE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED143072

USES AND ABUSES OF ADVERSARY EVALUATION:

A CONSUMER'S GUIDE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Blaine R Worthen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

Blaine R. Worthen
Northwest Regional Educational Laboratory

and

W. Todd Rogers
University of British Columbia

March 1977

Paper presented at the annual meeting of the American Educational Research Association, New York City, April 5, 1977, as part of a symposium entitled "The Adversary Evaluation Model: A Second Look."

EA 009 775

USES AND ABUSES OF ADVERSARY EVALUATION: A CONSUMER'S GUIDE

Even in the infant field of educational evaluation, the so-called adversary model of evaluation stands out as a relative newcomer. Guba (1965) suggested over a decade ago that educational evaluation might well adopt or adapt aspects of the legal paradigm. Apparently it was five years later before the first adversary evaluation in education was conducted (Owens, 1971). Since then, only a handful of evaluators have either conducted adversary evaluations or written about them. In a recent informal survey, Owens and Hiscox (1977) identified only six evaluations which they judged to be truly adversarial in nature (Owens, 1971; Hiscox and Owens, 1975; Wolf, 1975; Stenzel, 1976; Levine, 1976; and Northwest Regional Educational Laboratory, 1977).[1] Even if one included other studies which might be viewed as adversary evaluations (e.g., Stake and Gjerde, 1975; Kourilsky and Baker, 1976), it would seem safe to state that no more than eight or ten such evaluations have been conducted throughout the nation. Add the thoughtful conceptual work on adversary evaluation (e.g., Owens, 1972; Wolf, 1973), and there is still a paucity of information about this widely publicized evaluation method. Much more thought and experience is necessary before it will be clear whether the adversary method has the potential claimed by its proponents. Hopefully this symposium will help broaden the dialogue about adversary evaluation and lead to more careful analysis and experimentation in this area.

---

[1]There are evaluations using multiple advocate teams, each of which develops an independent position which may or may not be in opposition to other teams' positions (e.g., Reinhard, 1974). These are not included in this paper as examples of adversary evaluation.

The Focus of this Paper

When this symposium was planned, others were to argue for and against the adversary approach, while our assignment was to take the more neutral ground and provide an objective analysis of situations and settings where adversary evaluation would be beneficial and where it would be ill-advised. In the interim since that planning, however, we have both been involved in a large-scale adversary evaluation, one as a member of one of the opposing evaluation teams and one as co-director of the overall study and arbiter in disputes and negotiations between the teams.[2] Although we both began by believing the adversary approach would be very useful for that particular evaluation, the difficulties we experienced as we conducted the study were severe enough to give us serious second thoughts about the whole business of "adversarying." In fact, in the midst of the evaluation, we were sufficiently disenchanted that it was tempting to agree with Popham and Carlson (1977) in condemning the whole approach.

The presentation of the 3 on 2 final adversary reports and the "aftermath" of the evaluation have changed our perspective considerably, however. We have also been influenced by reactions of many key people affected by the evaluation. Although we still have reservations and cautions to share with you, we are convinced that the basic concept of adversary evaluation has real merit, if it is applied with prudence and judgment to those situations where it would be both

[2]This evaluation, which serves as the basis for much of the experience reported in this paper, is the Northwest Regional Educational Laboratory (NWREL) evaluation of the Hawaii 3 on 2 program, a large, controversial statewide team teaching program in the primary grades (NWREL, 1977). This study will hereafter be referred to as the 3 on 2 evaluation.

appropriate and advantageous. There is clearly real potential in the adversary approach for making evaluation findings more meaningful to educational decision makers. This does not mean that we accept without reservation all the claims made by proponents of adversary evaluation (e.g., Wolf, 1975; Wright and Sachse, 1977). We are frankly fearful that overly zealous supporters may fail to be sufficiently introspective to find and correct critical flaws in the adversary concept. We are equally fearful that preoccupation with the paraphernalia of the adversary model could cause evaluators to overlook the real benefits and problems that can result from its use.

In the remainder of this paper, we will present and discuss nine issues or questions which we think are central to the future of adversary evaluation in education. In stating our position on each issue, we hope to generate productive dialogue which can lead to development and refinement in the use of adversary evaluation.

1. Is There a Clearly Delineated Adversary Model of Evaluation Which Evaluators or Decision Makers Can Apply?

One of the authors has argued elsewhere (Worthen, 1977) that the term "evaluation models" is a misnomer when applied to the current conceptualizations about educational evaluation. This argument, which will not be repeated here, is not intended to denigrate the largely helpful suggestions which arise in the literature, but only to correctly describe them for what they are and are not. In no instance is the term model less appropriate than in the case of the so-called adversary model of evaluation. None of the criteria for models stated by Kaplan (1964) or other philosophers of science is met; adversary evaluation offers no unified framework or coherent set of principals. It is only a rubric under which to describe a collection of divergent approaches

which might loosely be referred to as adversarial in nature. In its broad sense, the term refers to all evaluations where there is planned opposition in the points of view of different evaluators or evaluation teams. The Websterian sense of "contending with, opposing" is central to this general definition. The fact that an evaluation approach includes a planned effort to generate opposing points of view within the overall evaluation is the sine qua non here, whereas the labeling becomes less important.

As Owens and Hiscox's (1977) descriptions make abundantly clear, none of the prior adversary evaluations (or writings on which they are based, are sufficiently well developed to set a standard for future efforts or to serve as a model of even the specific adversary approach employed. As yet there is little beyond personal preference to determine whether adversary hearings, debates or other approaches might be best in specific evaluation settings. Each approach should be further developed, applied in varied educational contexts, and studied to determine its relative utility under varying conditions. Given sufficient experience, Darwinian principles might apply and result in one specific adversary method proving best for most educational evaluations. In the meantime, it seems most defensible to use the term "adversary evaluation" in a broad sense and avoid the artificially precise and misleading terminology of "evaluation model."

The remainder of this paper assumes the notion of planned opposition among evaluators to be the only requirement for adversary evaluation.[3]

[3]The full range of forms this might take must await further development. In the meantime, it is obvious that some of the discussion in this paper will apply more directly to one type of adversary evaluation than another. We will leave it for others to tease out those specific applications.

2. Is the Legal Paradigm the Best Approach to Adversary Evaluation in Education?

Much of the effort to apply adversary evaluation in education has drawn on courtroom procedures, with an advocate and an adversary questioning and cross-examining witnesses and applying rules of admissibility of evidence customary in legal proceedings. If you were to ask any ten educational evaluators to describe the adversary evaluation approach, nine would probably talk in terms of witnesses, cross-examination, the jury system, and so forth.

The legal paradigm has intriguing possibilities for some evaluation situations, and Wolf (1973) has provided a good analysis of certain of these. We are not inclined, however, to view the legal paradigm as necessarily the best pattern for adversary evaluation. We tend to agree with Levine (1974) in favoring adversary evaluation more as a broad philosophical orientation which may be expressed in many forms. For example, cross-examination and juries may be appropriate in applying the courtroom model to educational evaluation, but they are hardly essential to conducting an adversary evaluation. In fact, one of our greatest concerns is that evaluators will seize on some of the more trivial features of the courtroom and fail to isolate and extract those adversarial aspects which might be most pertinent in educational evaluation. All that we have read and seen suggests to us that rigid adherence to the legal perspective is likely to result in weak adversary evaluations and an eventual rejection of the whole concept.

It might be useful to illustrate a few aspects of the legal system which seem to us unnecessary or downright inappropriate in educational evaluation.

First, we believe some of our colleagues should be chided for their compulsion to replicate even the theatrical aspects of the courtroom in

their adversary hearings. Cloaking the person presiding over an educational hearing in a black robe seems as pretentious and inane as placing powdered wigs on senators presiding over congressional hearings.

Second, we believe use of the legal model can result in a seductive slide into what might be termed an "indictment mentality," which can do a disservice both to evaluation efforts and to the programs being evaluated. Adversary evaluation literature which invokes the legal model tends to use terms such as "statement of charges" (Hiscox and Owens, 1975), "guilty or not guilty" (Levine, 1976), and the like. That orientation may be appropriate when there is a formal complaint against an educational program, as in the recent investigation of the University of Massachusetts School of Education programs. But formal complaints, plaintiffs and litigants are conspicuously absent in the typical educational evaluation--and rightly so. Evaluation in education should aspire to be an instrument for improving educational programs, not for determining their guilt or innocence. Although it is true that evaluators must of necessity render judgments of worth, that seems to us a far cry from invoking a model in which the program stands as "accused" on specific charges.

It is not just the vocabulary of the legal model that is problematic, but its characteristic of serving only when there is a problem to be solved. There is already too much of a tendency to view evaluation as something you do when a program is in trouble, when there is a crisis or failing which requires correction. It would be unfortunate if this tendency were exacerbated and evaluations conducted only when a complaint has been lodged, an accusation leveled, an offending program accused. It is precisely this orientation which

we fear may be a side effect of basing evaluations on the legal model, or for that fact, on any model which is meant to be applied only in problem-solving or crisis situations. It would be far more salutary if educators came to view evaluation as something which was routinely carried out to help them keep their programs operating at maximum effectiveness and efficiency. If advocates of the judicial approach respond that they only intend legal concepts to be applied to adversary evaluations which are conducted where complaints and charges are involved, many of the above concerns would be eased.

Obviously, one should not dismiss all aspects of the legal paradigm as inappropriate. For example, cross-examination (properly conducted) would seem to have a potentially useful role in evaluations which use human testimony as a major source of data. Of course one can use cross-examination by adversaries without requiring full or even partial courtroom procedures. Witness congressional hearings or interviews conducted jointly by partisan interviewers. Wolf (1975) and Hiscox and Owens (1975) have shown that one can adapt portions of the legal model without adopting it in its entirety.[4]

Hiscox and Owens (1975, p. 8) list five advantages which could accrue from lessening adherence to a strict legal model in adversary evaluations. Briefly, they are: (1) adversary evaluations could be conducted with lower investments of time and money; (2) adversary evaluations would be less dependent on availability of trained legal professionals; (3) adversary hearings

[4]Even here we believe there needs to be more attention to developing an adversary evaluation approach which would be suitable in education, for routine non-problem settings, without straining too hard to bend to our use an extant approach which is built on assumptions and for situations markedly different from those that apply in most educational evaluations.

could be more easily understood by evaluators and decision makers; (4) greater flexibility in addressing non-dichotomous issues would result; and (5) adversary evaluations or reporting could be conducted without formal hearings. We agree with these points. Although we believe the legal paradigm has merit as a heuristic, we also feel it carries many features which could be detrimental to educational evaluations. We hope others will view it with appropriate skepticism and entertain other alternatives before deciding which adversary approach is most suited to their needs in educational evaluation.

3. Does Adversary Evaluation Provide Decision Makers with the Full Range of Information Needed to Make Informed Decisions?

During our adversary evaluation of the Hawaii 3 on 2 Program, we worried considerably about whether the strong pro and con positions which were taken might increase the probability that an extreme decision might be made without due consideration of the full range of possible decisions which might be made. Would adversary evaluation result in an unwitting loss of the middle ground? In the typical evaluation, where an evaluator is charged with strict neutrality and objectivity, the middle ground might well receive as much attention as the ends of the spectrum. But what about adversary proceedings where the antagonists anchor the ends of the decision spectrum and choose to ignore the middle? Which best serves the decision maker, conflict or compromise, contrast or convergence, polarized positions or plea-bargaining? Does the adversary approach lend itself to the type of diagnostic information which is so often needed by the thoughtful decision maker?

Wrestling with these questions forced us to examine them in terms of three other questions: (1) does adversary evaluation provide a solution to the problem

of evaluator's biases slipping unnoticed into the evaluation; (2) is there a possibility for convergence in adversary evaluation; and (3) should an effort be made to present equally strong positive and negative arguments in adversary evaluations? Each of these areas is discussed briefly below.

Adversary Evaluation and Evaluator's Biases. Proponents of adversary evaluation (e.g., Wright & Sachse, 1977) have argued that evaluators are not the impartial, objective paragons they purport to be, and that they bring with them certain biases, often unrecognized, that influence their findings. Adversary evaluations are proposed as a solution since they intentionally counter-balance biases. One evaluator (or team) is assigned to present the positive case and is expected to be biased in favor of the program, while another is expected to be opposed to the program and be biased against it. The object then is not elimination of bias but rather balancing bias and making it public. Of course, still other biases and predispositions of the evaluators are unlikely to be affected by the mere assignment to a position. An individual evaluator's biases will obviously influence the rigor with which he can defend or criticize a program. Imagine the plight of Ralph Nader if he were assigned to defend a program or product. There is no great insight here, merely a reminder that bias is not magically eliminated or rendered inoperable by efforts to balance it.

Convergence in Adversary Evaluation. In moving from the usual evaluation stance of neutrality to that of having two biased protagonists, educators stand both to gain and lose. The gain is likely to be an increase in the spectrum of data and interpretations provided to decision makers; few other evaluation approaches seem likely to push as far in both the

positive and negative directions as the adversary method. The loss could easily come from unnecessary polarization that shifts attention away from the middle ground so often essential to rational decision making.

Many adversary supporters (e.g., Hiscox and Owens, 1975) have claimed that conclusions and recommendations agreed to by both sides may be held with greater confidence by a decision maker. While this seems patently sensible, experience with adversary evaluations suggests such agreement is unlikely to be a spontaneous by-product of the sparring and jousting that often occurs between adversaries. Most adversary approaches have a competitive element; one of the adversaries will probably win and the other lose. When competition is high, cooperation tends to be lower. There is less of an inclination to search for agreement than is true under more collaborative circumstances. In highly competitive evaluations, mutual agreements are often abandoned in the adversaries' rush to dispute each datum in hopes of turning it to their own advantage. When "winning" is at stake, even "black is black" pronouncements are sometimes questioned by seemingly rational opponents. Antagonists are often leery of agreements, even about things they may both believe, especially if they construe the agreement as potentially injurious to their case(s). Shared conclusions in adversary evaluation are not easy to come by. Most adversary approaches could profit from a better mechanism for seeking and reporting areas of agreement.

In the Hawaii 3 on 2 evaluation, it was decided that presenting the strongest possible pro and con cases would best serve the needs of the decision makers. As the evaluation progressed, it became apparent that

the evaluation teams were trending toward "all or none" recommendations--maintain the program in its entirety or eliminate it completely. With that posture, it became difficult to get either team very "psyched up" about evaluation approaches which ferreted out features of the program which could be jettisoned without loss, or features which should be retained even if the overall program were scrapped. Several members of the evaluation team worried that this approach would result in loss of important diagnostic information which did not support either extreme position.[5] The polarized report was very well received in Hawaii, however, and only two board members or administrators complained about the fact that the evaluation would lead to either a "go" or "no go" decision. Although we believe the 3 on 2 evaluation was a good one, it would be much better in our judgment had some way been found to converge on areas where both teams agreed there were strengths or weaknesses.

There is some empirical evidence which bears on the reconciliation of views between evaluators. Kourilsky and Baker (1976) reported that college students produce significantly better evaluations of a project when using an adversary approach than under two other approaches which do not involve confrontation. Their adversary treatment required that adversaries reconcile their views and produce a single recommendation to the decision maker. This method was found to produce significantly superior results over other less adversarial methods. Unfortunately, the study did not include a comparison

[5]Much such information was included in the technical evaluation report, but since it was not presented in the more provocative adversary reports, it seems operationally to have had little impact on subsequent decisions about the program.

treatment in which adversaries were not asked to converge, so one pivotal bit of data is still lacking.

Relative Strength of Adversary Positions. It may not be an explicit assumption of the approach, but many adversary evaluations proceed as if there is an unspoken obligation to present two equally convincing cases, one pro and one con. Of course no one would tolerate an adversary who slacked and presented a weaker case than was deserved on the basis of the data; but what about the advocate who errs in the other direction, who feels compelled to keep up with the opposition, even if it means straining or ignoring the data? Here is where we part philosophical company with some colleagues who seem to sincerely believe that a program is not represented well unless both sides are argued equally convincingly. That orientation strikes us as appropriate in a forensic society where the result of the debate seldom impacts on the proposition, but not in an evaluation where the outcome will influence real programs and real people.

Like the legal paradigm, the debate model also carries with it many irrelevancies that should be strained out before the model is applied to education. The critical difference is in the fact that the touchstones of debate are polemics and persuasion, not truth, which is central to the validity of evaluation studies. Debates surely use facts and cannot normally afford to ignore them, at least not totally. But seldom is the debater forced to adhere as tightly to the plain unadorned facts as is the conscionable evaluator. Logic can provide a permissive climate for manipulating the data until its form is favorable. Probably more sophistry results from debaters' perversions of syllogistic logic than any other self-deception known to man. At least par:

of this tendency must be traced to efforts to build strong cases on flimsy foundations.

Our recommendation in this area would be for decision makers to think carefully about the charge they give to adversary evaluators. We believe the appropriate mandate is that of presenting the most positive and most negative cases possible on the basis of the evidence which exists. Within that framework, the evaluator might be encouraged to employ all the techniques of persuasion possible, just so adversarial zeal does not lead to flights of fancy or specious arguments that exceed the evidence. Of course, one could depend on rebuttals or cross-examination to expose fallacies and errors introduced by overly enthusiastic adversaries, but that dependence seems optimistic. It would be better to require documentation and evidence for arguments at the outset rather than to allow unsubstantiated assertions to become part of the substance that is contested in an adversary evaluation. If such mandates to evaluators were made clear, then no evaluator would feel compelled to fabricate a strong positive case when none exists, where the overwhelming weight of the evidence reveals the product or program to be without redeeming features, or vice versa.

4. Does Challenging of Evidence in Adversary Evaluations Reduce Their Credibility?

Data in a typical evaluation are only contested by outside critics, usually after the fact. In adversary evaluations, the data themselves can become a source of dispute between adversaries, and this has both pluses and minuses. For example, one can argue (at least in educational evaluation) that all data and the instruments and designs that produce them are open to some degree of question. Therefore, they may as well be questioned by opponents within the

adversary framework as by possibly less informed persons at a later point in time. There is also a potential, however, that disputes among adversaries over the validity of data will, at best, shift attention away from the substance of the evaluation to its processes, and, at worst, will jeopardize the credibility of the evaluation. Imagine an evaluation with two major types of data, let us say test scores and observer ratings. Imagine that one adversary makes every effort to discredit test scores (which not coincidentally favors the opposition) while the other seriously questions the observer ratings. If both evaluators are skillful at pointing out and perhaps dramatically magnifying the flaws which exist in most data collection techniques, the net result could well be to discredit the entire data base and destroy the credibility of the study. For example, Popham and Carlson (1977) stated their view that the arbiters in the 3 on 2 evaluation exercised good judgment in allowing their team to argue that the tests used in the evaluation were invalid. Perhaps, but at least some in the Hawaii State Board of Education felt differently. When asked in a recent questionnaire, "Does the advocate-adversary approach provide decision makers with the evidence they need to make a choice?", one board member wrote, "Not when the integrity of the evaluation instruments is attacked. That attack on the instruments completely destroyed the credibility of the study's overall findings and, in a politically charged issue, allowed board members to ignore the evaluation and do whatever they wanted." That may be an over-reaction, but it does demonstrate the risks of allowing opponents to extend their contentions to the data base. In the heat of competition, methodological pimples have a way of getting portrayed as terminal illnesses.

Now this should not be construed as a suggestion that bad instruments or data should be tolerated. The point is simply that techniques should be built into adversary evaluations to produce a common core of data that both sides would accept as valid for purposes of judging the program. Variables and the best methods for measuring them should be agreed upon in advance, not determined on the basis of partisanship. Surely evidence should be challenged and only the most solid used as a basis for evaluative judgments, but it would seem wise to deal with this issue early in an adversary evaluation so the focus in the final stages can be on inferences, arguments and judging the program rather than quarreling about the adequacy of the evidential basis for the evaluation. We all enjoy the cleverness of the defense attorney who holds up an optometrist's chart at the back of the courtroom to prove the prosecution witness is myopic and could not possibly have identified the defendant at the distance claimed. High drama should be reserved for the TV courtroom; in educational evaluations, such faulty witnesses should be dispensed with much earlier and not at the final report stage.

Considerably more thought must be given in this area to working out rules for judging admissibility and validity of evidence in adversary evaluations.

5. In What Settings and Under What Circumstances Would an Adversary Evaluation be Appropriate?

Even the most enthusiastic advocate of adversary evaluation is unlikely to argue that the approach would be appropriate in any evaluation. In an effort to get others' opinions on this issue, a questionnaire was developed[6] and sent to key figures in Hawaii, both decision makers and evaluators. They were

[6]This questionnaire was developed jointly by one of the authors and William J. Wright.

asked when they thought it would be appropriate to use an adversary evaluation. A majority of the respondents indicated that they saw an adversary evaluation as appropriate in the following instances:

a. When the program is controversial and people are polarized in their opinions over it (93%).

b. When decisions must be made about whether to continue or terminate a program (81%).

c. When the program is large and affects many people (77%).

d. When there are many different audiences for the evaluation report (65%), and

e. When the evaluation is conducted by persons external to the program (56%).

Very few respondents felt adversary evaluation would be appropriate when the evaluation was conducted by internal evaluators (15%) for purposes of making decisions about how to improve the program (15%).[7]

These reactions and our own biases lead us to suggest several factors which we think should govern decisions about when to use the adversary approach.[8]

The Decision. As implied above, adversary evaluation would seem less relevant for most formative decisions than for summative decisions about program continuation. Using adversary evaluation also assumes the full range of decision alternatives is available to the decision maker. Aside from

---

[7]Written comments suggested some respondents reacted this way because of their perception that adversary evaluations give up the diagnostic middle ground relevant to program improvement decisions.

[8]Obviously these points may need to be altered somewhat if one chooses to look at a specific type of adversary approach, such as the legal or debate model.

the intellectual enjoyment, arguing from adversary positions is of dubious worth if one side has no chance, e.g., lack of funds dictates that a program be terminated regardless of its quality. If clearly competing courses of action are not available, the adversary approach has little to recommend it.

The Object of the Evaluation. The Hawaii respondents felt adversary evaluation was most appropriate for large, controversial programs which had a variety of interested audiences. We tend to agree. The adversary approach is an ambitious, costly and sometimes cumbersome method. As such, it should be reserved for cases which warrant the increased investment of time and money and where its use would add significantly to the results of the study. It would be presumptuous for us to suggest types of programs where it should be used, but it seems clear that one does not wheel out heavy artillery for every minor skirmish.

Clarity of Issues. Adversary evaluation loses its meaning unless issues to be addressed by the adversaries are clearly identified and adhered to. If one adversary dwells solely on test scores and the other deals exclusively with financial aspects of the program, the potential advantages of the adversary method are seriously diluted.

Credibility. There are instances where a program is so controversial that no evaluation of it will be believed unless it can be shown unequivocally that the evaluation made every effort to represent fairly both sides of the issue. This is often true where previous evaluations of the program have been condemned as one-sided or discounted on grounds of evaluator bias. Here the adversary approach comes into its own with its built-in neutrality (or balanced bias) which allows both sides of an issue to be well illuminated.

A related feature of the adversary approach is its potential for diffusing political heat surrounding an evaluation. Some evaluators have privately proposed that the best place for this approach might be the "hot potato" evaluations where the evaluator will be pilloried no matter which way the results come out. As one wag put it, "It's hard to claim an evaluation is wrong when it argues both sides of the issue." There may be some truth in that bit of facetiousness, since the 3 on 2 evaluation was conducted in a political inferno and not only survived but was generally acclaimed in wide press coverage as "unbiased," "a comprehensive study," and a "balanced evaluation." At least no one claimed that the evaluation was biased, and the heated exchanges and dual recommendations provided all the fodder necessary for the administrative and political decision makers. The evaluators did not get drawn back into the fray to defend recommendations which were under attack. Those recommendations had already been attacked within the evaluation. Not that the evaluation was not criticized--one legislator went so far as to print an attack against both sides of the evaluation, for using "disembodied statistics" and tests that would have received higher scores from the "intoxicated and tightly controlled students of Nazi Germany." Yes, even the adversary approach fails to dispel some folks' distrust of anything as anti-humanitarian as a test item.

Courageous Clients. By now it should be apparent that not all administrators are likely to have the heart to initiate adversary evaluations of their programs. Hiscox and Owens (1975, p. 6) found that

> ". . . some administrators indicated that they would not be interested in using an adversary hearing as a decision-making tool. They felt that many of their decisions were based largely on personal experience,

were not to be resolved publicly or had political overtones so that a logical decision based on solicited facts might not be adequate for their needs."

Such administrators probably would prefer avoiding any evaluation at all. They should be doubly tempted to avoid an adversary evaluation, where few stones remain unturned. The competitive nature and processes of adversary proceedings also make them less predictable than more standard approaches. An administrator who willingly requests an adversary evaluation is either a self-confident and visionary leader or uninformed about the approach.

Costs. Much of the informal dialogue among evaluators the past two or three years has questioned whether the adversary approach is worth its considerable costs. In fairness, the cost depends on how you play the game. If a full-blown courtroom procedure is employed, the cost is likely to be proportionately quite high for the amount of data produced. If every phase of an evaluation involves the kind of two-party cross-checking Wright and Sachse (1977) describe, the resultant data may be better, but you need not be a mathematical whiz to predict the cost will double. Even a simple debate as a vehicle for presenting findings from a standard evaluation is an added expense.

The real question, however, is not cost but cost-effectiveness or cost-benefit. On these dimensions, it seems apparent that benefit must be argued on grounds that adversary evaluation increases things like representativeness of the data, fairness of the instruments, communication between evaluators and decision makers, and identification of all the pros and cons. Whether adversary evaluation really provides more benefits will remain an open question until someone sees fit to research the issue. In the interim, the survey of Hawaii educators is provocative. When asked if they felt the

information produced by the adversary approach was worth the cost of having two teams involved, 78 percent said it was, and another 15 percent said it was worth more than the cost. Only 7 percent felt it was not worth the money. Of course, these reactions should not be generalized beyond the evaluation to which they were reacting, but they do demonstrate that even a relatively costly adversary evaluation can be viewed as worth the cost.[9]

6. How Should Adversary Evaluations be Conducted?

This paper will not address this issue satisfactorily, for there must be at least as many answers as there are different approaches one might make to adversary evaluation. The best method for conducting an evaluation with two independent adversary teams with separate budgets is obviously different from that for using a debate as an enlightening way to present a standard evaluation report. Rather than speculate on how to conduct such variations, it might be helpful to list some critical features in the 3 on 2 evaluation since it represents one view of how an adversary evaluation might be structured.

First, two evenly balanced teams were formed.

Second, both teams worked together to create the best possible design and choose the best possible instruments to provide the common core of data to be shared by both teams. The intent here was to develop a comprehensive data base that would be accepted as valid by both teams.[10] The thought was to get all persons to think about information needed by both adversaries before they knew which position they would represent.

---

[9] Also, it might be reinterated that the basic adversary concept can be implemented without the heavy costs associated with some of the approaches discussed earlier.

[10] The fact that this intent was not realized, and one team dhose to attack the tests does not negate the usefulness of this point. It merely underscores the need for clearer and firmer ground rules from the outset.

Third, teams were assigned to adversary positions for the balance of the evaluation. Data collection and analysis were mostly joint efforts, with checks and balances built in to prevent either team from influencing the outcomes in their favor. Reporting was decidedly adversarial with a written and live debate format, buttressed by a neutral technical report.

Wright and Sachse (1977) have described several phases of evaluation during which adversary input is useful. In our view, the adversary approach reaches its zenith in the reporting stage. Much of the positive reception to the Hawaii 3 on 2 evaluation is probably attributable to the report format. The interest and positive effect stimulated by the public debate format must be viewed as considerable since the findings of our study generally parallel those of the previous evaluation which was soundly censured two years earlier.

There are probably many sensible approaches which could still take advantage of the adversary report format, while streamlining the process and cutting costs. For example, one individual or team could conduct the entire evaluation, with two outsiders assigned to present the advocate and adversary cases from the data generated by the evaluation. The same outsiders could also obviously be called in earlier to ensure balance in the choice of variables and instruments, check the design for fairness, and so forth.

Other possibilities are left to the imagination of the reader.

7. Do Adequate Guidelines Exist for Use in Managing Adversary Evaluations?

Given the newness of adversary evaluation, it would be foolish to expect adequate guidelines to have emerged for any of the variations which have been proposed. Some extant paradigms (e.g., debate or courtroom models) do

have carefully prescribed operating guidelines, but they have to be bent so far to fit educational evaluations that they become largely inapplicable. Other de novo approaches which have been developed have even fewer procedural guidelines to suggest.

To further dialogue in this area, a few administrative or planning guidelines which our experience and observation suggest would be relevant in any adversary evaluation are listed below.

First, we share Popham's view that the director must be concerned with achieving as much balance as possible in the relevant skills and strengths of the adversaries.

Second, it is paramount that the evaluation ground rules which will apply in the study be spelled out in specific detail before the evaluation begins. Such ground rules must be in place and agreed to by all parties prior to the time that partisan positions are assigned. Decisions about admissibility and validity of evidence should be agreed to early and adhered to throughout. Decide early on the role of the judge or the arbiters. What criteria constitutes an objection that should be sustained? What rules govern how far arbiters can go in insisting that all arguments that draw on the data be adequately documented, or that claims not supported by data be removed? Sufficient attention given to spelling out such ground rules adequately at the outset of an adversary evaluation will avoid many problems later on. Do not assume that general ground rules will suffice and that usual collegial congeniality will make compromise and resolution simple in areas overlooked in initial guidelines. In our experience, it seems unrealistic to expect such behavior in a confrontive methodology calculated to create opposition.

The need for specific guidelines for each evaluation is important in view of the absence of a body of procedural canons and guidelines. Attention to this step at the outset may help to solve Popham and Carlson's concern over the absence of appellate mechanisms for adversary evaluations.

Another area in need of careful management in adversary evaluations involves decisions about data to be collected. Kourlisky and Baker (1976) noted that adversary evaluation results in longer evaluation reports and requires the development of guidelines for careful but parsimonious reporting. The Hawaii 3 on 2 experience corroborated theirs, for adversary team members tended to collect or request a good deal of data without adequate plans for its use. Part of the problem was an apparent reluctance to allow their opponents to get ahead in the data-aggregation game. The result was that a fair amount of data was underutilized.

There are many other guidelines that might be suggested, but brevity requires that we quickly move to our final two main points.

8. Does Adversarial Evaluation Alter the Nature of Evaluation Ethics?

The field of educational evaluation does not yet have an articulated, formalized code of ethics. The work of the committee on evaluation standards empaneled jointly by AERA and other professional associations (Stufflebeam, 1977) is directly relevant, but it is still too early to tell just how much guidance that effort will give in the area of ethical practices. In the interim, there does seem to be general agreement among most evaluators on certain minimum essentials of ethical behavior, and at the heart of these lies venerable principles such as impartiality and neutrality.

In the typical educational evaluation, evaluators are to be neutral and impartial, leading to a fair unprejudiced evaluation. The evaluator's role is roughly parallel to that of a judge, where impartiality is the sine qua non in weighing the evidence.[11] In adversary evaluation, it is only the overall structure and process which is obliged to be impartial. The evaluators are intentionally partisan and their roles approximate that of lawyers, where neutrality gives way to advocacy.

Theoretically, shifting from non-adversary to adversary evaluation does not lessen the impartiality with which decisions and judgments will be made. But it most decidedly forces the individual evaluator to put aside reverence for personal impartiality and adopt standards of behavior more like those followed by debaters and attorneys.[12] Is that good? Or is the sudden shift to a new role disruptive and dysfunctional for evaluators? Frankly, we have no idea. It seems unlikely that evaluators' psyches will be permanently damaged by occasional forays into fields where different standards are followed. But might adversarial behavior prove addictive, making the tough job of remaining unprejudiced in non-adversary evaluations even tougher for the evaluator who traverses the ethical boundary too frequently? Time will tell.

Although not ethical considerations per se, it is interesting to note the behavioral modifications which are sometimes wrought by adversary contests.

---

[11] According to Wright and Sachse, such impartiality may exist more in folklore than fact. We agree that evaluators are fallible, but are unconvinced that they fail to be impartial as frequently as our colleagues' rhetoric implies.

[12] In this post-Watergate period where potshots at attorneys are a lamentable national pastime, we stress that our reference here is to personal impartiality, not personal integrity.

Win/lose situations can tax one's professionality and it is a credit to the Hawaii 3 on 2 evaluation team leaders that throughout the 3 on 2 debates they refrained from challenging one another's integrity and ancestry. Of course, that might have been partly due to the rule we were forced to make early in the evaluation that adversaries could not make disparaging remarks about one another's mothers.

Adversary evaluation also provides memorable moments, like hearing a hard-nosed empiricist whimsically scold his opponent for presenting a "data-drenched report." Another was watching newsmen scurry for the telephones when one adversary referred to the program as "a beautiful dream from 1968 that remained only a beautiful dream eight years later." (That bit of prose made great copy in all the Honolulu papers.) And then there was the emotional moment when one adversary asked for special indulgence from the "jury" because he was small and bald. If nothing else, the adversary approach is hardly boring.

9. Are Educational Evaluators Competent to Conduct Advocacy Evaluations?

This issue cannot be addressed well until someone completes a careful analysis of the skills and knowledge required of evaluations in the various adversary approaches. In the meantime, predicting who will make good, adversary evaluations must be categorized with other forms of crystal ball gazing. There are basic considerations like technical ability, communications skills, and general ability, to be sure, but those are too gross to be very helpful. Lack of information about what skills are needed also makes it difficult to develop criteria to measure how well an adversary evaluator is performing. Adding adversary skills to the repertoire of techniques provided

in training programs for educational evaluators will also be impeded until better information is available.

In the meantime, based on our limited observations of adversary evaluations, we suspect that most educational evaluators are not well prepared to play the adversarial role, especially if the legal model is adopted. Hiscox (1976) noted the following problems in previous adversary hearings:

1. Political and professional considerations make it difficult for educator-advocates to attack incompetence of statements and evaluations presented as evidence.

2. Educator-advocates fail to appear adversarial; they often make points for the other side with their questions and/or evidence.

3. People unskilled in soliciting "testimony" often get rambling, unproductive evidence.

We have noticed similar tendencies in educators we have watched function in adversary hearings based loosely on the legal model. Questioning skills were notably lacking and witnesses were permitted to ramble in long monologs that addressed the questions indirectly, if at all. Probing of obvious contradictions in testimony usually stopped short of highlighting the contradiction, as if the most important thing were to avoid embarrassing the witness. The relevance of testimony to major issues in the case was often left obscure.

Educators may be able to function more readily in the debate model, but even this bit of optimism is mostly speculative. If adversary evaluation is to become a potent force in educational evaluation, more thought must be given to defining and providing training in this area.

Conclusion

We have discussed what we believe to be some major potentials and pitfalls of adversary evaluation. We have expressed our suspicion that the courtroom

model may have limited utility for adversary evaluations in education, and we have pointed out difficulties that seem inherent in the debate model. We have argued that the existence of opposing viewpoints is the core of adversary evaluation, not adherence to existing formats for presenting contrasting views. We have suggested that educational evaluators might develop more appropriate adversary methods tailored specifically for the field of education. We have addressed nine issues which should be considered by anyone intending to use the adversary approach.

Where our analysis has been critical, it is prompted by a desire to see improvements in an approach which we feel could be very useful in selected evaluation settings. Adversary evaluation seems to hold considerable promise for improving the data base on which important educational decisions are made--if the pitfalls we have outlined can be resolved.

REFERENCES

Guba, E. G. Evaluation in field studies. Address at evaluation conference sponsored by the Ohio State Department of Education, Columbus, Ohio, 1965.

Hiscox, M. D., & Owens, T. R. Attempts at implementing an educational adversary model. Paper proposed for presentation at Third Annual Pacific Northwest Educational Research and Evaluation Conference, Seattle, 1975.

Hiscox, M. D. Some of the problems of previous adversary evaluations. Northwest Regional Educational Laboratory. Mimeo, 1976.

Kaplan, A. The conduct of inquiry. San Francisco: Chandler, 1964.

Kourilsky, M., & Baker, E. L. An experimental comparison of interaction, advocacy and adversary evaluation. CeDR Quarterly, Summer 1976.

Levine, M. Scientific method and the adversary model, some preliminary thoughts. American Psychologist, September 1974.

Levine, M. Experiences in adapting the jury trial to the problem of educational program evaluation. Unpublished manuscript, State University of New York at Buffalo, 1976.

Northwest Regional Educational Laboratory. 3 on 2 evaluation report, 1976-77. Volume I, Technical Report. Northwest Regional Educational Laboratory, January 1977.

Owens, T. R. Application of adversary proceedings for educational evaluation and decision making. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1971.

Owens, T. R. Educational evaluation by adversary proceeding. In E. R. House (ed.), School evaluation: the politics and the process. Berkeley: McCutchan, 1973.

Owens, T. R., & Hiscox, M. D. Alternative models for adversary evaluation: Variations on a theme. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.

Popham, W. J., & Carlson, D. Deep dark deficits of the adversary evaluation model. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.

Reinhard, D. L. Methodology development for advocate team use for input evaluation. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1973.

Stake, R. E., & Gjerde, C. An evaluation of TCITY: The Twin City Institute for Talented Youth. Evaluation Center, College of Education, Western Michigan University, Report #1. November 1975.

Stenzel, N. Meta-evaluation of the IPS adversary hearing evaluation. Illinois Department of Education. Mimeo, 1976.

Stufflebeam, D. L. An overview of the Joint Committee's project on evaluation standards. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.

Wolf, R. L. Application of select legal concepts to educational evaluation. Unpublished doctoral dissertation, University of Illinois, 1973.

Wolf, R. L. Trial by jury: A new evaluation method, 1. The process. Phi Delta Kappan, November 1975.

Worthen, B. R. Eclecticism and evaluation models: Snapshots of an elephant's anatomy. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.

Wright, W. J., & Sachse, T. Payoffs of adversary evaluation. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1977.